

Syntactic Structures of Regular Languages

Ondřej Klíma and Libor Polák *

Department of Mathematics and Statistics, Masaryk University
Kotlářská 2, 602 00 Brno, Czech Republic
{klima,polak}@math.muni.cz
<http://www.math.muni.cz>

Abstract. We introduce here the notion of syntactic lattice algebra which is an analogy of the syntactic monoid and of the syntactic semiring. We present a unified approach to get those three structures.

1 Introduction

The Eilenberg like theorems establish bijections between the class of all (in some sense generalized) varieties of regular languages and the class of all pseudovarieties of certain algebraic structures — see Eilenberg [3], Pin [6], Straubing [11], Polák [9]. The classical result concerns the varieties of regular languages and pseudovarieties of finite monoids. The goal is an algorithmic procedure for deciding the membership of a given language in various significant classes of regular languages. A basic source of that theory is the book by Pin [8].

The aim of the present contribution is to introduce modifications of the notion of the syntactic monoid which would be useful in other variants of Eilenberg type theorems. As well-known the syntactic monoid of a language L over the alphabet A can be viewed as the transformation monoid of the minimal complete deterministic automaton \mathcal{A}_L of L . More precisely, we let words of A^* to act on states of \mathcal{A}_L and the composition of such transformations corresponds to multiplication in the syntactic monoid.

To get analogues of A^* with the multiplication we consider structures with more operations, namely we use here the following three term algebras:

- F is the absolutely free algebra over the alphabet A with the operation symbol \cdot and nullary symbol λ ,
- to get F' we enrich the previous signature by binary \wedge and nullary 0 ,
- to get F'' we enrich the last signature by binary \vee and nullary ∇ .

Now we let to act our terms on the set 2^{A^*} of all languages over A in a natural way (the formal definitions are in Section 3).

We show here that identifying terms of F (F' and F'') giving the same transformations we get exactly the free monoid A^* over A , (the free semiring A^\square over A and the free so-called lattice algebra A^\diamond over A), respectively. Informally,

* The authors were supported by the Institute for Theoretical Computer Science (GAP202/12/G061), Czech Science Foundation.

all our considerations concern three worlds: world of monoids – the classical one (Pin [7,8]), world of semirings (considered also in Polák [9,10]), and that of lattice algebras – a new contribution.

When generating subalgebras in 2^{A^*} by a single regular language L using elements of those algebras of transformations and choosing the final states appropriately, we get the classical minimal complete deterministic finite automaton of L (here called the canonical finite automaton of L), the canonical meet automaton of L , see Section 6 of Polák [9], and the canonical lattice automaton of L , respectively. Then transforming those automata accordingly we get the corresponding syntactic structures.

Our construction are also accompanied by examples. Finally, we present some instances of future Eilenberg type theorems.

2 Preliminaries

By an *idempotent semiring* we mean the structure $(S, +, \cdot, 0, 1)$ where $(S, +, 0)$ is a commutative idempotent monoid with the neutral element 0, $(S, \cdot, 1)$ is a monoid with the neutral element 1 and the zero element 0, and the operations $+$ and \cdot satisfy the usual distributivity laws. The set S can be ordered by $a \leq b$ iff $a + b = a$ ($a, b \in S$). Then 0 becomes the greatest element.

The elements of the free idempotent semiring A^\square over the set A can be represented by finite subsets of A^* . This representation is one-to-one. Operations are the operation of union and the obvious multiplication, \emptyset is the neutral element for $+$, the zero for \cdot and $\{\lambda\}$ is the neutral element for \cdot . The interpretation of $U = \{u_1, \dots, u_k\}$, $k \geq 0$, $u_1, \dots, u_k \in A^*$ is $u_1 \wedge \dots \wedge u_k$.

Next structure is the free bounded distributive lattice A^\diamond over A^* . Its elements can be represented as

$$\{\{u_{1,1}, \dots, u_{1,r_1}\}, \dots, \{u_{k,1}, \dots, u_{k,r_k}\}\},$$

$$k, r_1, \dots, r_k \geq 0, u_{i,j} \in A^* \text{ for } i = 1, \dots, k, j = 1, \dots, r_i.$$

Taking the inner sets incomparable we get the unique representatives. The interpretation of the above element is $(u_{1,1} \wedge \dots \wedge u_{1,r_1}) \vee \dots \vee (u_{k,1} \wedge \dots \wedge u_{k,r_k})$. See Grätzer [2]. Notice that $\{\emptyset\}$ is the greatest element and \emptyset is the smallest one.

The structure A^\diamond is equipped also with a multiplication, namely extend the multiplication from A^* to A^\diamond using

$$\mathcal{U} \cdot (\mathcal{V} \wedge \mathcal{W}) = \mathcal{U} \cdot \mathcal{V} \wedge \mathcal{U} \cdot \mathcal{W}, (\mathcal{U} \wedge \mathcal{V}) \cdot w = \mathcal{U} \cdot w \wedge \mathcal{V} \cdot w \quad (\diamond\diamond)$$

for $\mathcal{U}, \mathcal{V}, \mathcal{W} \in A^\diamond$, $w \in A^*$. Similarly for the dual operation \vee .

If we use symbols \mathcal{U}, \mathcal{V} for elements from A^\diamond , we assume that

$$\mathcal{U} = \{U_1, \dots, U_k\}, U_1 = \{u_{1,1}, \dots, u_{1,r_1}\}, \dots, U_k = \{u_{k,1}, \dots, u_{k,r_k}\},$$

$$\mathcal{V} = \{V_1, \dots, V_\ell\}, V_1 = \{v_{1,1}, \dots, v_{1,s_1}\}, \dots, V_\ell = \{v_{\ell,1}, \dots, v_{\ell,s_\ell}\},$$

where $k, r_1, \dots, r_k, \ell, s_1, \dots, s_\ell \geq 0$, and all $u_{i,j}$'s and $v_{i,j}$'s are from A^* .

We consider various kinds of automata. All of them are deterministic and complete and could have infinite number of states. When using the term *semi-automata*, no initial nor final states are specified.

Having an equivalence relation ρ on a set G and an element $a \in G$, we denote by $a\rho$ the class of G/ρ containing a .

3 Transformation structures

Let A be a finite non-empty set. We consider the actions of term algebras mentioned above on languages over the alphabet A .

Monoids. Let F be the absolutely free algebra (that is, the algebra of all terms) over A with respect to the binary operational symbol \cdot and nullary operational symbol λ .

We define inductively the actions of elements of F on subsets of A^* :

$$L \circ \lambda = L, \quad L \circ a = a^{-1}L \text{ for } a \in A, \quad L \circ (u \cdot v) = (L \circ u) \circ v \text{ for } u, v \in F. \quad (*)$$

For $u, v \in F$, we put $u \rho^* v$ iff $(\forall L \subseteq A^*) L \circ u = L \circ v$.

Proposition 1. *The relation ρ^* is a congruence relation on F and F/ρ^* is isomorphic to the free monoid A^* over A via the extension of the mapping $a\rho^* \mapsto a$, $a \in A$.*

Proof. Let $u, v, w \in F$. If $u \rho^* v$ then, for each $L \subseteq A^*$, we have $L \circ u = L \circ v$. Therefore $L \circ (u \cdot w) = (L \circ u) \circ w = (L \circ v) \circ w = L \circ (v \cdot w)$, which gives $u \cdot w \rho^* v \cdot w$. Similarly, $(L \circ w) \circ u = (L \circ w) \circ v$ and $L \circ (w \cdot u) = (L \circ w) \circ u = (L \circ w) \circ v = L \circ (w \cdot v)$, which gives $w \cdot u \rho^* w \cdot v$. Thus ρ^* is a congruence relation on F .

Now we prove that, for each $u, v, w \in F$, we have $(u \cdot v) \cdot w \rho^* u \cdot (v \cdot w)$, $\lambda \cdot u \rho^* u$ and $u \cdot \lambda \rho^* u$. Indeed, choosing $L \subseteq A^*$, it holds $L \circ ((u \cdot v) \cdot w) = (L \circ (u \cdot v)) \circ w = ((L \circ u) \circ v) \circ w = (L \circ u) \circ (v \cdot w) = L \circ (u \cdot (v \cdot w))$. Further, $L \circ (\lambda \cdot u) = (L \circ \lambda) \circ u = L \circ u$, and $L \circ (u \cdot \lambda) = (L \circ u) \circ \lambda = L \circ u$.

Thus we can omit brackets in elements of F and λ acts as a neutral element. Therefore every element of F/ρ^* can be represented by a word from A^* . It remains to show that different words u and v represent different elements of F/ρ^* . Indeed, for $u \neq v$, we have $1 \in \{u\} \circ u$ but $1 \notin \{u\} \circ v$. \square

The structure $(2^{A^*}, A, \circ)$ is called here the *canonical semiautomaton* on A . Due to Proposition 1, given a regular language L over A , we can generate by L in $(2^{A^*}, A, \circ)$ the *canonical finite semiautomaton* of L ; namely $(\{L \circ u \mid u \in A^*\}, A, \circ)$ – it is really finite due to Proposition 2. Notice that $L \circ u = u^{-1}L$ for all $u \in A^*$. Taking L as the unique initial state and $T = \{L \circ u \mid 1 \in L \circ u\}$ as the set of all final states, we get the *canonical finite automaton* of L . This is the famous Brzozowski's construction of the minimal complete automaton for L – see [1]. For the sake of completeness we prove a part of that result.

Proposition 2 ([1]). *Given a regular language L over the alphabet A , the automaton $(\{u^{-1}L \mid u \in A^*\}, A, \circ, L, T)$ is finite and accepts L .*

Proof. Take a finite automaton (Q, A, \cdot, i, T') accepting L . For $q \in Q$, put $L_q = \{v \in A^* \mid q \cdot v \in T'\}$. In particular $L_i = L$. Notice that, for each $q \in Q$ and $a \in A$, we have $L_{q \cdot a} = a^{-1}L_q$. Furthermore, for each $u \in A^*$, we have $u^{-1}L = L_{i \cdot u}$ and therefore there are only finitely many $(u^{-1}L)$'s.

Now the automaton $(\{u^{-1}L \mid u \in A^*\}, A, \circ, L, T)$ accepts a word u if and only if $1 \in L \circ u = u^{-1}L$, that is $u \in L$. \square

Semirings. Let F' be the absolutely free algebra over A with respect to the operational symbols \cdot, λ , binary symbol \wedge and nullary symbol 0 . We define inductively the actions of elements of F' on 2^{A^*} : we use the formulas from $(*)$ for $u, v \in F'$ and

$$L \circ 0 = A^*, \quad L \circ (u \wedge v) = (L \circ u) \cap (L \circ v) \text{ for } u, v \in F'. \quad (\square)$$

For $u, v \in F'$, we put $u \rho^\square v$ iff $(\forall L \subseteq A^*) L \circ u = L \circ v$.

Proposition 3. *The relation ρ^\square is a congruence relation on F' and F'/ρ^\square is isomorphic to the free idempotent semiring A^\square over A via the extension of the mapping $a\rho^\square \mapsto a$, $a \in A$.*

Proof. Let $u, v, w \in F'$. If $u \rho^\square v$ then, for each $L \subseteq A^*$, we have $L \circ u = L \circ v$. We get $u \cdot w \rho^\square v \cdot w$ and $w \cdot u \rho^\square w \cdot v$ as in the case of Proposition 1.

Further, $L \circ (u \wedge w) = (L \circ u) \cap (L \circ w) = (L \circ v) \cap (L \circ w) = L \circ (v \wedge w)$, which gives $u \wedge w \rho^\square v \wedge w$. In the same way we can prove that $w \wedge u \rho^\square w \wedge v$. Thus ρ^\square is a congruence relation on F' .

Now we show that $(F'/\rho^\square, \wedge, 0\rho^\square)$ is an commutative idempotent monoid with the neutral element $0\rho^\square$. The commutativity and associativity of \wedge is clear as the fact that $0\rho^\square$ is a neutral element for the operation \wedge . To show the idempotency of \wedge notice that, for each $L \subseteq A^*$ and $u \in F'$, we have $L \circ (u \wedge u) = (L \circ u) \cap (L \circ u) = L \circ u$.

The proof of the associativity of \cdot on F'/ρ^\square and the fact that $\lambda\rho^\square$ is a neutral element for the operation \cdot is similar to that for monoids. The fact that $0\rho^\square$ is a zero element for \cdot is clear.

Finally, we prove the distributivity laws. Let $L \subseteq A^*$, $u, v, w \in F'$. Then $L \circ (u \cdot (v \wedge w)) = (L \circ u) \circ (v \wedge w) = (L \circ u) \circ v \cap (L \circ u) \circ w = L \circ u \cdot v \cap L \circ u \cdot w = L \circ (u \cdot v \wedge u \cdot w)$. Similarly, $L \circ ((u \wedge v) \cdot w) = (L \circ (u \wedge v)) \circ w = (L \circ u \cap L \circ v) \circ w = (L \circ u) \circ w \cap (L \circ v) \circ w = (L \circ u \cdot w) \cap (L \circ v \cdot w) = L \circ (u \cdot w \wedge v \cdot w)$.

We have proved that F'/ρ^\square is an idempotent semiring. Therefore every element of F'/ρ^\square can be represented by $u_1 \wedge \dots \wedge u_k$ with $k \geq 0$ and $u_1, \dots, u_k \in A^*$. To get the unique representation of such element we use the idempotency and commutativity law and represent the element in F'/ρ^\square by the set $\{u_1, \dots, u_k\}$. Having such two different sets $\{u_1, \dots, u_k\}$ and $\{v_1, \dots, v_\ell\}$ ($\ell \geq 0, v_1, \dots, v_\ell \in A^*$), we show that they are not ρ^\square -related. Indeed, put $L = \{u_1, \dots, u_k\}$. Then $\lambda \in L \circ \{u_1, \dots, u_k\} = u_1^{-1}L \cap \dots \cap u_k^{-1}L$ and $\lambda \in L \circ \{v_1, \dots, v_\ell\}$ would give $\{v_1, \dots, v_\ell\} \subseteq \{u_1, \dots, u_k\}$. Take $L = \{v_1, \dots, v_\ell\}$ in this case. \square

The structure $(2^{A^*}, A, \circ, \cap)$ forms the *canonical meet semiautomaton* on A . Moreover, given a regular language L over A , we can generate by L in $(2^{A^*}, A, \circ, \cap)$ the *canonical finite meet semiautomaton* of L ; namely $(\{L \circ U \mid U \in A^\square\}, A, \circ, \cap)$. Taking L as the unique initial state and all states containing 1 as the set of all final states, we get the *canonical finite meet automaton* of L .

Lattice algebras. Let F'' be the absolutely free algebra over A with respect to the operational symbols $\cdot, \lambda, \wedge, 0$, binary \vee and nullary ∇ . We use $(*)$, (\square) with $u, v \in F''$ and

$$L \circ \nabla = \emptyset, \quad L \circ (u \vee v) = (L \circ u) \cup (L \circ v) \text{ for } u, v \in F'' . \quad (\diamond)$$

For $u, v \in F''$, we put $u \rho^\diamond v$ iff $(\forall L \subseteq A^*) L \circ u = L \circ v$.

Proposition 4. *The relation ρ^\diamond is a congruence relation on F'' and F''/ρ^\diamond is isomorphic to the free distributive lattice A^\diamond over A^* with (\diamond) (see Section 2) and their duals via the extension of the mapping $a\rho^\diamond \mapsto a$, $a \in A$.*

Proof. Let $u, v, w \in F''$. If $u \rho^\diamond v$ then, for each $L \subseteq A^*$, we have $L \circ u = L \circ v$. We get $u \cdot w \rho^\diamond v \cdot w$, $w \cdot u \rho^\diamond w \cdot v$, $u \wedge w \rho^\diamond v \wedge w$, $w \wedge u \rho^\diamond w \wedge v$ as in the case of Proposition 3. Furthermore, $L \circ (u \vee w) = (L \circ u) \cup (L \circ w) = (L \circ v) \cup (L \circ w) = L \circ (v \vee w)$, which gives $u \vee w \rho^\diamond v \vee w$. In the same way we can prove that $w \vee u \rho^\diamond w \vee v$. Thus ρ^\diamond is a congruence relation on F'' .

Now we state the properties of operations $\wedge, \vee, \cdot, 0, \nabla$ and λ on F''/ρ^\diamond . Proofs of all statements are straightforward and omitted. The operation \wedge is commutative, associative and idempotent, 0 is the neutral element and ∇ is the zero. The operation \vee is commutative, associative and idempotent, ∇ is the neutral element and 0 is the zero. The operations \wedge and \vee are connected by the distributivity laws. The operation \cdot is associative, λ is the neutral element, 0 and ∇ are right zeros, and $0 \cdot a \rho^\diamond 0$, $\nabla \cdot a \rho^\diamond \nabla$ for all $a \in A$. Finally, the distributivity $u \cdot (v \vee w) = u \cdot v \vee u \cdot w$ holds for arbitrary $u, v, w \in F''$ and the distributivity $(u \vee v) \cdot w = u \cdot w \vee v \cdot w$ for $u, v \in F''$ and $w \in A^*$. Similarly for the operation \wedge .

We have proved that every element of F''/ρ^\diamond can be represented as $(u_{1,1} \wedge \dots \wedge u_{1,r_1}) \vee \dots \vee (u_{k,1} \wedge \dots \wedge u_{k,r_k})$, where $k, r_1, \dots, r_k \geq 0$ and $u_{i,j} \in A^*$ for all $i = 1, \dots, k$, $j = 1, \dots, r_i$. (Here $k = 0$ corresponds to the element ∇ and $k = 1, r_1 = 0$ corresponds to the element 0 .) Using the idempotency and commutativity of \wedge and \vee we can write such element even as $\{\{u_{1,1}, \dots, u_{1,r_1}\}, \dots, \{u_{k,1}, \dots, u_{k,r_k}\}\}$. To get canonical forms remove the richer one from each pair of comparable inner sets. We show that such different sets \mathcal{U} and \mathcal{V} represent elements of F'' which are not ρ^\diamond -related. If $U_i \notin \mathcal{V}$, take $L = U_i$. Then $1 \in L \circ \mathcal{U}$ and $1 \in L \circ \mathcal{V}$ would give that $V_j \subseteq U_i$ for some $V_j \in \mathcal{V}$ and we can take $L = V_j$. Therefore F''/ρ^\diamond is isomorphic to A^\diamond . \square

Example 1. The distributivity (\diamond) is not true for $w \in A^\diamond$ in general. Indeed, let $a, b \in A$ be different and let $L = \{aa, bb\}$. Then $\lambda \in L \circ (a \cdot (a \vee b) \wedge b \cdot (a \vee b))$ but $L \circ ((a \wedge b) \cdot (a \vee b)) = \emptyset$.

The structure $(2^{A^*}, A, \circ, \cap, \cup)$ forms the *canonical lattice semiautomaton* on A . Moreover, given a regular language L over A , we can generate by L in $(2^{A^*}, A, \circ, \cap, \cup)$ the *canonical finite lattice semiautomaton* of L ; namely

$$(\{L \circ \mathcal{U} \mid \mathcal{U} \in A^\circ\}, A, \circ, \cap, \cup).$$

This structure is already mentioned in Klíma [4]. Taking L as the unique initial state and all states containing 1 as the set of all final states, we get the *canonical finite lattice automaton* of L .

4 An example

To show examples of three types of the canonical finite automata, we consider the language $L = a^+b^+$ over the alphabet $A = \{a, b\}$. At first, in the canonical finite automaton \mathcal{A}_L , we have four states $L = a^+b^+$, $K = a^{-1}L = a^*b^+$, $b^{-1}L = \emptyset$ and $b^{-1}K = b^*$. There is just one final state containing the empty word, namely the state b^* . The automaton is depicted on Figure 1.

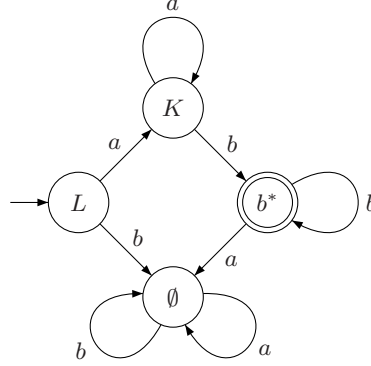


Fig. 1. The canonical finite automaton of the language $L = a^+b^+$.

To construct the canonical finite meet automaton \mathcal{M}_L of the same language $L = a^+b^+$ we need to consider all possible intersections of states from \mathcal{A}_L . There are two new states: the intersection $K \cap b^* = b^+$ and the intersection of the empty system $\bigcap_{\emptyset} = A^*$. The canonical finite meet automaton is depicted on Figure 2.

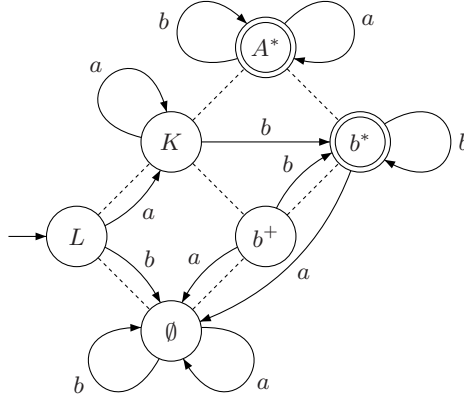


Fig. 2. The canonical finite meet automaton of the language $L = a^+b^+$.

Dashed lines indicate the inclusion relation on the set of all states. The inclusion relation completely describes a semilattice structure of the meet automaton \mathcal{M}_L .

Finally, we consider the canonical finite lattice automaton \mathcal{L}_L of the language $L = a^+b^+$, which is depicted on Figure 3. There is only one new state, which is $K^\lambda = K \cup b^* = K \cup \{\lambda\}$ in addition to the canonical finite meet automaton \mathcal{M}_L . Now, the inclusion relation describes a lattice structure of \mathcal{L}_L .

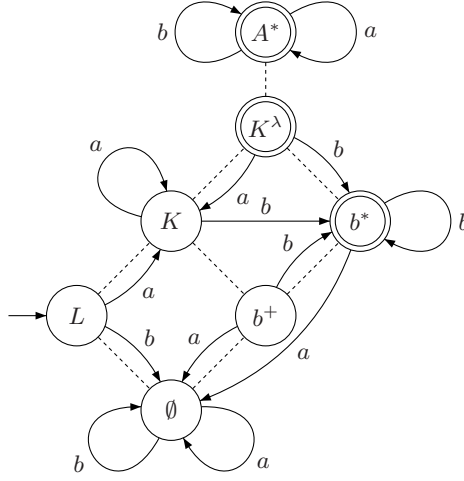


Fig. 3. The canonical finite lattice automaton of the language $L = a^+b^+$.

5 Syntactic structures

The basic tool of the algebraic language theory is the concept of the syntactic monoid of a regular language. It is a certain finite quotient of the free monoid on the corresponding alphabet. We recall here its definition and its construction. Then we consider modifications for the remaining two worlds.

Monoids. Given a regular language L over the alphabet A , we define the so-called *syntactic congruence* \sim_L^* of L on A^* as follows: for $u, v \in A^*$, put

$$u \sim_L^* v \text{ iff } (\forall p, q \in A^*) (puq \in L \iff pvq \in L).$$

The following is a folklore result.

Proposition 5. *The relation \sim_L^* is a congruence relation on A^* . Moreover, for $u, v \in A^*$, it holds that $u \sim_L^* v$ iff $(\forall p \in A^*) p^{-1}L \circ u = p^{-1}L \circ v$. So the structure A^*/\sim_L^* is isomorphic to the transformation monoid of the canonical finite semiautomaton of L .*

Proof. Clearly, the relation \sim_L^* is reflexive, symmetric and transitive. Furthermore, for $u, v, w \in A^*$, if $u \sim_L^* v$ then $uw \sim_L^* vw$ and $wu \sim_L^* wv$. Clearly, the fact $u \sim_L^* v$ is equivalent to $(\forall p, q \in A^*) (q \in (pu)^{-1}L \iff q \in (pv)^{-1}L)$, which is $(\forall p \in A^*) (pu)^{-1}L = (pv)^{-1}L$, that is $(\forall p \in A^*) p^{-1}L \circ u = p^{-1}L \circ v$. \square

The structure A^*/\sim_L^* is called the *syntactic monoid* of L .

Semirings. Given a regular language L over the alphabet A , we define the so-called *syntactic (semiring) congruence* \sim_L^\square of L on A^\square as follows: For $U = \{u_1, \dots, u_k\}$, $V = \{v_1, \dots, v_\ell\} \in A^\square$, we put $U \sim_L^\square V$ iff

$$(\forall p, q \in A^*) (pu_1q \in L, \dots, pu_kq \in L \iff pv_1q \in L, \dots, pv_\ell q \in L).$$

Proposition 6 ([10]). *The relation \sim_L^\square is a congruence relation on A^\square . Moreover, for $U, V \in A^\square$, it holds that $U \sim_L^\square V$ iff $(\forall p \in A^*) p^{-1}L \circ U = p^{-1}L \circ V$. So the structure $A^\square / \sim_L^\square$ is isomorphic to the transformation semiring of the canonical finite meet semiautomata of L .*

Proof. To show that the relation \sim_L^\square is a congruence relation on A^\square is easy and similar to the case of monoids. Clearly, the fact $U \sim_L^\square V$ is equivalent to $(\forall p, q \in A^*) q \in (pu_1)^{-1}L \cap \dots \cap (pu_k)^{-1}L \iff q \in (pv_1)^{-1}L \cap \dots \cap (pv_\ell)^{-1}L$. The last formula can be written as $(\forall p \in A^*) (pu_1)^{-1}L \cap \dots \cap (pu_k)^{-1}L = (pv_1)^{-1}L \cap \dots \cap (pv_\ell)^{-1}L$, which is $(\forall p \in A^*) p^{-1}L \circ U = p^{-1}L \circ V$. \square

The structure $A^\square / \sim_L^\square$ is called the *syntactic semiring* of L .

The examples of syntactic semirings can be found e.g. in [9]. In the paper [10] it is described how one can compute the syntactic semiring algorithmically from the syntactic monoid. For the hand made computations we can use Proposition 6. For example, in the case of the language $L = a^+b^+$, we can choose the words λ , a , b , ab , and ba to represent five different transformations. There are no others, because both a and b are idempotent elements of both syntactic monoid and semiring and ba is a zero element. Moreover, ba is the minimum element in the syntactic semiring, because ba transforms all states, with exception of A^* , to the state \emptyset . So, if we want to compute all elements of the syntactic semiring, it is enough to consider only intersections of the elements λ , a , b and ab . The crucial observation is that both $\lambda \wedge ab$ and $a \wedge b$ give the same transformation as the intersection of any triple of elements. Hence in the syntactic semiring there are, besides the element 0 and elements λ , a , b , ab , and ba , just five elements given by intersections $\lambda \wedge a$, $\lambda \wedge b$, $\lambda \wedge ab$, $a \wedge ab$ and $b \wedge ab$. In Table 1 we present how all these elements transform the canonical finite meet automaton. The semilattice part of the syntactic semiring is fully described by Figure 4.

	L	K	b^*	b^+	A^*	\emptyset
λ	L	K	b^*	b^+	A^*	\emptyset
a	K	K	\emptyset	\emptyset	A^*	\emptyset
b	\emptyset	b^*	b^*	b^*	A^*	\emptyset
ab	b^*	b^*	\emptyset	\emptyset	A^*	\emptyset
ba	\emptyset	\emptyset	\emptyset	\emptyset	A^*	\emptyset
0	A^*	A^*	A^*	A^*	A^*	A^*
$\lambda \wedge a$	L	K	\emptyset	\emptyset	A^*	\emptyset
$\lambda \wedge b$	\emptyset	b^+	b^*	b^+	A^*	\emptyset
$\lambda \wedge ab$	\emptyset	b^+	\emptyset	\emptyset	A^*	\emptyset
$a \wedge ab$	b^+	b^+	\emptyset	\emptyset	A^*	\emptyset
$b \wedge ab$	\emptyset	b^*	\emptyset	\emptyset	A^*	\emptyset

Table 1. The transformations of \mathcal{M}_L for the language $L = a^+b^+$.

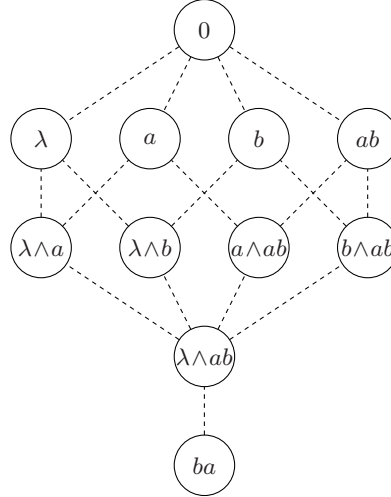


Fig. 4. The semilattice order of the syntactic semiring of the language $L = a^+b^+$.

Lattice algebras. Given a regular language L over the alphabet A , we define the so-called *syntactic (lattice) congruence* \sim_L^\diamond of L on A^\diamond as follows: for $\mathcal{U}, \mathcal{V} \in A^\diamond$ we put $\mathcal{U} \sim_L^\diamond \mathcal{V}$ iff $(\forall p, q \in A^*)$

$$(pU_1q \subseteq L \text{ or } \dots \text{ or } pU_kq \subseteq L \iff pV_1q \subseteq L \text{ or } \dots \text{ or } pV_\ell q \subseteq L).$$

Proposition 7. *The relation \sim_L^\diamond is a congruence relation on A^\diamond . Moreover, for $\mathcal{U}, \mathcal{V} \in A^\diamond$, it holds that $\mathcal{U} \sim_L^\diamond \mathcal{V}$ iff $(\forall p \in A^*) p^{-1}L \circ \mathcal{U} = p^{-1}L \circ \mathcal{V}$. So*

the structure $A^\diamond / \sim_L^\diamond$ of L is isomorphic to the algebra of transformation of the canonical finite lattice semiautomata of L .

Proof. To show that the relation \sim_L^\diamond is a congruence relation on A^\diamond is easy and similar to the case of monoids.

Clearly, $\mathcal{U} \sim_L^\diamond \mathcal{V}$ is equivalent to $(\forall p, q \in A^*)$

$$\begin{aligned} q &\in ((pu_{1,1})^{-1}L \cap \dots \cap (pu_{1,r_1})^{-1}L) \cup \dots \cup ((pu_{k,1})^{-1}L \cap \dots \cap (pu_{k,r_k})^{-1}L) \\ \iff q &\in (pv_{1,1})^{-1}L \cap \dots \cap (pv_{1,s_1})^{-1}L \cup \dots \cup (pv_{\ell,1})^{-1}L \cap \dots \cap (pv_{\ell,s_\ell})^{-1}L, \end{aligned}$$

which is

$$\begin{aligned} (\forall p \in A^*) &((pu_{1,1})^{-1}L \cap \dots \cap (pu_{1,r_1})^{-1}L) \cup \dots \cup ((pu_{k,1})^{-1}L \cap \dots \cap (pu_{k,r_k})^{-1}L) \\ &= ((pv_{1,1})^{-1}L \cap \dots \cap (pv_{1,s_1})^{-1}L) \cup \dots \cup ((pv_{\ell,1})^{-1}L \cap \dots \cap (pv_{\ell,s_\ell})^{-1}L). \end{aligned}$$

that is $(\forall p \in A^*) p^{-1}L \circ \mathcal{U} = p^{-1}L \circ \mathcal{V}$. \square

The structure $A^\diamond / \sim_L^\diamond$ is called the *syntactic lattice algebra* of L .

Due to the space limitations, we are not able to present here the syntactic lattice algebra of the language $L = a^+b^+$. We mention here just the following aspect. In the syntactic lattice algebra of L the elements given by $\lambda \wedge ab$ and by $a \wedge b$ are different. Indeed, $K^\lambda \circ (\lambda \wedge ab) = b^*$ and $K^\lambda \circ (a \wedge b) = b^+$. Interestingly, these two terms $\lambda \wedge ab$ and $a \wedge b$ give the same element of the syntactic semiring. This example shows, that the syntactic lattice algebra can not be viewed as an extension of the syntactic semiring by adding joins.

6 General algebras

The Eilenberg like theorems establish bijections between certain varieties of regular languages and pseudovarieties of certain algebraic systems. Not every finite monoid is isomorphic to a syntactic one, therefore we have to generate the appropriate pseudovariety. Similarly in remaining words.

Monoids. One considers varieties of languages and pseudovarieties of finite monoids. The Eilenberg theorem can be find in [7].

Semirings. One considers the so-called conjunctive varieties and pseudovarieties of finite semirings. For more details see e.g. [9].

Lattice algebras. The following new definition of a notion of lattice algebras is a part of an effort of formulation of Eilenberg like theorem using the notion of syntactic lattice algebra. Such a theorem is not formulated or even proved in this paper. On the other hand Section 3 and 5 should validate the concept of the following formalism. In that definition we try to follow the basic properties of the structure F''/ρ^\diamond which is isomorphic to A^\diamond by Proposition 7. In particular, we must take into account that some distributivity laws have certain values restricted to words, which must be viewed as products of generators. Simply

speaking, the finitely generated lattice algebras should be exactly homomorphic images of the free lattice algebras over appropriate finite alphabet.

A *lattice algebra* is 8-tuple $(K, \wedge, \vee, \cdot, P, \perp, \top, 1)$ where (K, \wedge, \vee) is a bounded distributive lattice with the bottom element \perp and the top element \top , $(K, \cdot, 1)$ is a monoid with right zero elements \perp and \top , P is a finite subset of K such that the lattice (K, \wedge, \vee) is generated by the set of all products of elements from P and $\top \cdot p$ and $\perp \cdot p$ hold for $p \in P$, and finally such that the distributivities

$$q \cdot (r \wedge s) = q \cdot r \wedge q \cdot s, \quad q \cdot (r \vee s) = q \cdot r \vee q \cdot s,$$

$$(q \wedge r) \cdot p = q \cdot p \wedge r \cdot p, \quad (q \vee r) \cdot p = q \cdot p \vee r \cdot p$$

hold for $q, r, s \in K$ and $p \in P$.

Notice that, considering A^\diamond , take P equal to the image of A , \top the image of $\{\emptyset\}$ and \perp the image of \emptyset .

7 Examples

We show an example of a class of regular languages which can be characterized by the properties of their syntactic semirings. The class consisting of complements of these languages was mentioned in [5], where a similar characterization was given, however in a slightly different way.

Example 2. We consider the class \mathcal{C} consisting from all regular languages for which syntactic semirings have the binary operations \cdot and \wedge equal on all elements different from 0. In other words, the syntactic semirings satisfy the equation $x \cdot y = x \wedge y$ for all $x \neq 0 \neq y$. We could mention here that the case when the operations \cdot and \wedge are completely equal means that the zero element 0 for \cdot , which is neutral element for \wedge is equal to the neutral element λ for \cdot . Notice that the zero element is equal to the neutral element only in trivial monoids.

We claim that a language L over the alphabet A belongs to \mathcal{C} if and only if $L = B^*$ for some $B \subseteq A$ or $L = \emptyset$.

Let $\mathcal{M}_L = (Q_L, A, \circ, \cap)$ be the canonical finite meet semiautomaton of a regular language $L \in \mathcal{C}$, where $Q_L = \{L \circ U \mid U \in A^*\}$. Since the syntactic semiring of L satisfies $x \cdot y = x \wedge y$ for all $x \neq 0 \neq y$, the meet semiautomaton \mathcal{M}_L satisfies

$$(\forall q \in Q_L, u, v \in A^*) \quad q \circ uv = q \circ u \cap q \circ v.$$

It follows that

$$(\forall q \in Q_L, a_1, a_2, \dots, a_n \in A) \quad q \circ a_1 \dots a_n = q \circ a_1 \cap q \circ a_2 \cap \dots \cap q \circ a_n.$$

Now for $u = a_1 \dots a_n$, with $n > 0$, $a_1, \dots, a_n \in A$, we have $u \in L \iff 1 \in L \circ u \iff (\forall i = 1, \dots, n) 1 \in L \circ a_i \iff a_1, a_2, \dots, a_n \in L$. If we denote $B = L \cap A$, then we can deduce that $L = B^+$ or $L = B^*$. If $B = \emptyset$, then both cases $L = \emptyset$ and $L = \emptyset^* = \{\lambda\}$ are possible. Now assume that $B \neq \emptyset$. If we take some $b \in L \cap A$, then $\lambda \in L \circ b = L \circ b \cap L \circ \lambda$ and we have $\lambda \in L$. Thus, in the case $B \neq \emptyset$, we can conclude that $L = B^*$.

Now we show the converse. The canonical finite meet semiautomaton and the syntactic semiring of $L = A^*$ are trivial. In the case $L = \emptyset$, the canonical finite meet semiautomaton has two elements \emptyset, A^* and the syntactic semiring consists from two transformations given by elements λ and 0 . For all $B \subseteq A$, $B \neq A$, the canonical finite meet semiautomaton of $L = B^*$ has just three states: B^*, A^*, \emptyset . Here we have three transformations: the constant transformation onto A^* , i.e. the transformation given by 0 ; the identical transformation given by words from $L = B^*$; and finally the transformation given by words outside B^* which maps both B^* and \emptyset onto \emptyset and A^* onto A^* . In all cases one can see that the binary operations \cdot and \wedge are equal on all elements different from 0 .

Notice that the class of all semirings satisfying this condition contains much more semirings than these syntactic semirings with one, two or three elements.

Now we show an easy example of a class of languages which can be characterized by properties of their syntactic lattice algebra.

Example 3. We consider the class \mathcal{C}' of all languages for which the syntactic lattice algebras satisfy the condition from the previous example, i.e. $x \cdot y = x \wedge y$ for all $x \neq 0 \neq y$, and in addition $a \vee b = \lambda$ for every pair of different letters $a \neq b$.

Since we can use the arguments from the previous example, we can assume that $L = B^*$ or $L = \emptyset$. In the later case the second condition is satisfied. Now in the first case $L = B^*$ the condition $B^* \circ a \cup B^* \circ b = B^*$, for $a \neq b$, is saying that $a \in B$ or $b \in B$. In other words $A \setminus B$ has at most one element. Thus languages over $A = \{a_1, \dots, a_n\}$ which belongs to \mathcal{C}' are exactly the languages $A^*, (A \setminus \{a_1\})^*, \dots, (A \setminus \{a_n\})^*$ and \emptyset .

Note that the class \mathcal{C}' has a natural counterpart consisting of the complements of the languages, namely $\emptyset, A^*a_1A^*, \dots, A^*a_nA^*, A^*$ for languages over $A = \{a_1, \dots, a_n\}$. Therefore, this class can be characterized by conditions valid in the syntactic lattice algebras $x \cdot y = x \vee y$ for all $x \neq \top \neq y$, and in addition $a \wedge b = \lambda$ for every pair of different letters $a \neq b$.

References

1. Brzozowski, J.: Derivatives of regular expressions. J. ACM 11, 481–494 (1964)
2. Grätzer, G.: *General Lattice Theory*, Second edition, Birkhäuser (2003)
3. Eilenberg, S.: *Automata, Languages and Machines*, Vol. B, Academic Press (1976)
4. Klíma, O.: On varieties of automata enriched with an algebraic structure (Extended abstract). In Proc. AFL 2014, EPTCS 151, 49–54 (2014), arXiv:1405.5272
5. Klíma, O., Polák, L.: Hierarchies of piecewise testable languages. In Proc. Developments Language Theory 2008, LNCS 5257, 479–490 (2008)
6. Pin, J.-E.: A variety theorem without complementation, *Izvestiya VUZ Matematika* **39** (1995) 80–90. English version: *Russian Mathem. (Iz. VUZ)* **39** (1995) 74–83
7. Pin, J.-E.: *Syntactic Semigroups*, Chapter 10 in *Handbook of Formal Languages*. G. Rozenberg and A. Salomaa eds, Springer (1997)
8. Pin, J.-E.: *Varieties of Formal Languages*. North Oxford Academic, Plenum (1986)

9. Polák, L.: Syntactic semiring of a language. In Proc. Mathematical Foundations of Computer Science 2001, LNCS 2136, 611–620 (2001)
10. Polák, L.: Syntactic semiring and universal automaton. In Proc. Developments Language Theory, Szeged 2003, LNCS 2710, 411–422 (2003)
11. Straubing, H.: On logical descriptions of regular languages, *Proc. LATIN 2002*, Springer Lecture Notes in Computer Science, Vol. 2286, pages 528–538 (2002)